Redefining Digital Neighborhoods: Filling the Gaps of Foursquare and Twitter

Cheng Yee Lim

October 23, 2016

Introduction

The old, stagnant delineation of neighborhoods can hardly capture the dynamic and fluid nature of neighborhoods. (Cranshaw et Al, 2011) The advent of the internet has enabled the use of social media data to describe city dynamics and outline digital neighborhoods. (Metz, 2012) However, previous papers which investigated digital neighborhoods have mainly focused on Foursquare and Twitter data, which often have time intervals between check-ins on these social media platforms. In this paper, we aim to fill this "gap" with Pokémon GO data, which are more fine-grained, as check-ins occur in closer intervals. With Pokémon GO, Foursquare and Twitter data, I revise the definitions of neighborhoods in cities with in Chicago, San Francisco and New York and compare the differences in digital neighborhoods outlined with each type of geo-located data.

Data and Methodology

Pokémon GO is a location-based augmented reality game, where players use their device's GPS to access real life Pokéstops and Pokémon Gyms. (Pokémon GO, 2016) Despite its recent release, Pokémon GO has 30 million daily active users in the United States, which is equivalent to 60% of Foursquare monthly users, and collects locational information when users check in at Pokéstops and Pokémon gyms. Pokémon GO data will be obtained from Niantic directly. Foursquare[1] data, containing user check-ins at a location, will be accessed with Foursquare application programming interfaces (API). (Foursquare, 2016) Furthermore, Twitter data will be gathered from Gnip, Twitter's enterprise API platform, and contains every single message as compared to the miniscule 1% sample from the public API. Twitter messages are geo-coded to the place where the individual posted their message. (Twitter, 2016)

The locations contained in the Foursquare, Twitter and Pokémon GO data are used as point inputs of the spatial database. These points are then aggregated based on the census block groups[2], forming the total count of digital events. However, simply counting the number of times people checked in to a location portrays a false indication of social media intensity as it fails to account for the underlying reference population of the census block. Therefore, one could use a density

[1] Foursquare is a location-based online social network that provides users a way to share their location with their friends by "checking in" to places they visit. (Foursquare, 2016)
[2] A census block group is the smallest geographical unit that the United States Census Bureau publishes sample data on.

measure instead, such as check-ins per capita, to reflect the true social media usage intensity. (Anselin and Williams, 2015) A common location quotient is as follows:

$$LQ = \frac{C_{bg} / C_{total}}{P_{bg} / P_{total}}$$

Where C refers to the number of check-ins, in the form of either tweets, foursquare check-ins or Pokémon GO Pokéstops and Pokémon gyms; P refers to the reference population from the US Census. The subscript bg and total pertains the measurements for the block group and the entire area respectively. As a result, census block groups with zero population, such as parks and open spaces, were removed as the check-ins cannot be interpreted in a meaningful way.

Thereafter, the local Moran statistic (Anselin, 1995) will be employed to identify clusters and spatial outliers, generating new boundaries for neighborhoods that can be used to make comparisons of different digital neighborhoods outlined. The differences between Twitter-based, Foursquare-based and Pokémon GO-based digital neighborhoods will help reveal the nature of check-ins on these platforms, and perhaps reveal insights on how to build weights to factor multiple digital data sources.

Suitability of Digital Data for Urban Sciences

With the rapid proliferation of smartphones and the emergence of location-based applications, there have been attempts to redefine neighborhoods by leveraging on digital information to account for the dynamic nature of urban areas. Digital data are often big and contain vast amounts of longitudinal information of many individuals. For instance, Twitter and Foursquare both have over 50 million monthly active users in the US and Pokémon GO has over 30 million daily active users. (Twitter, 2016) (Foursquare, 2016) (Niantec, 2016) Furthermore, the metadata of Twitter and Foursquare contain a multitude of information such as user information, content of the message, and latitude-longitude coordinates where the message is posted. The unprecedented amount of data can provide new insights on human activity spaces, mobility and digital inequality.

Past studies which leveraged on digital data to investigate neighborhoods focused on either Foursquare or Twitter data. (Cranshaw et Al, 2011) (Anselin and Williams, 2015) While the data collected were more continuous than before, it fell short of a near-continuity updating of locations as individuals lack the incentive to constantly update their Foursquare and Twitter accounts. On the other hand, users of Pokémon GO are incentivized to check in to Pokéstops to collect items for their games when they are travelling and are required to check in to Pokémon gym before they can use the play against other players. As a result, the locational information of Pokémon GO can provide the route users have taken with multiple Pokéstops checked in, filling up the “gaps” of Twitter and Foursquare data. However, there are more constraints on locations where users can check in, compared to Twitter and Foursquare, as locations of Pokéstops and gyms are fixed. There is also possibility of losing non-reactivity these digital observational data as users of Pokémon GO could change their usual travelling routes based on the game checkpoints.

Despite the flaws in the three types of digital data, these data are in fact complementary when aggregated and downplays the flaws in each dataset. Pokémon GO data reduces the omission of check-ins on Twitter and Foursquare check-ins as tweets are often posted more arbitrarily and Foursquare users tend to check-in at commercial, operational and residential venues. On the other hand, Foursquare and Twitter provide the possibility for users to post digital information at a location of their choice. Furthermore, the research question is one of the few studies that do not require a representative sample as it is precisely the difference in digital usage that we are studying. Thus, the use of these three digital sources, Foursquare, Twitter and Pokémon GO, do provide a good measure of digital intensity and a good starting point to study digital neighborhoods.

However, the use of longitudinal digital data is often susceptible to drifting due to changes in users, the way people use social media or games and systems. (Salganik, 2017) These drifts result in greater difficulty in studying long-term trends. For instance, system updates in Pokémon GO may introduce new Pokéstops, where individuals can check in at, resulting in a vast change in digital neighborhood boundaries. Usage drifts can also occur due to launching of new location-based services and remodeling of old services. Take Foursquare for example, Foursquare was changed to a local search service in summer 2014 and the check-in feature was moved to a companion application called Swarm. Therefore, it is imperative to monitor change in old services and changes in the location-based services landscape to update the model accordingly.

Feasibility

The presence of open API for many social media services has provided an exceptional opportunity to carry out empirical research based on large amounts of information. However, I hope to convince Niantic and Twitter to provide me with locational data of their users to conduct the research by providing marketing strategies for their firms with the results of the research. The use of location-based services to outline digital neighborhoods could highlight areas where the competitors have a stronger presence, which they have yet to penetrate. Furthermore, this paper is only done to progress urban science academic research and do not require contentious private information of their users.

Conclusion

In conclusion, the research question is well-framed in a way that leverages on the big, non-reactive and continuous nature of big data and reduces the common drawbacks such as incomplete and inaccessible data, unrepresentative samples and ethical contentions. This paper also serves as a good starting in evaluating the use of location-based augmented reality games data in future urban science research. Augmented reality games such as Pokémon GO could pose as potential data sources in complementing usual targets of geotagged data (i.e. Foursquare and Twitter) to mitigate the deliberate choices of users checking in at certain places on Foursquare and Twitter. This will enable better demarcations of digital neighborhoods and facilitate future research on computational urban sciences.

Bibliography

Anselin, L., & Williams, S. (2015). Digital neighborhoods. Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 1-24.

Anselin, L. 1995. "Local Indicators of Spatial Association – LISA." Geographical Analysis 27(2): 93–115.

Bruns, A., & Stieglitz, S. (2014). Twitter data: what do they represent?. it-Information Technology, 56(5), 240-245.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In International AAAI Conference on Weblogs and Social Media (p. 58).

Foursquare. (2016). Foursquarecom. Retrieved 24 October, 2016, from https://foursquare.com/about

Metz, R. (2012). MIT Technology Review. Retrieved 24 October, 2016, from https://www.technologyreview.com/s/427650/using-foursquare-data-to-redefine-a-neighborhood/

Pokemon GO. (2016). Pokemongocom. Retrieved 24 October, 2016, from http://www.pokemongo.com/en-us/

Salganik, Matthew J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press. Open Review Edition.

Twitter. (2016). Twittercom. Retrieved 24 October, 2016, from https://about.twitter.com/company